Semantics-informed geological maps: Conceptual modeling and knowledge encoding

Published version:

DOI:10.1016/j.cageo.2018.04.001

Semantics–informed geological maps: conceptual modeling and knowledge encoding

Vincenzo Lombardo[a,b], Fabrizio Piana[b], Dario Mimmo[c]

[a]*Università di Torino, CIRMA and Dipartimento di Informatica, Torino, Italy*
E-mail: vincenzo.lombardo@unito.it
[b]*CNR–IGG and Gi-RES srl, Torino, Italy*
E-mail: F.Piana@csg.to.cnr.it
[c]*Gi-RES srl and Università di Torino, Dipartimento di Informatica and Dipartimento di Scienze della Terra, Torino, Italy*
E-mail: dario.mimmo@gmail.com

Abstract

This paper introduces a novel, semantics-informed geologic mapping process, whose application domain is the production of a synthetic geologic map of a large administrative region. A number of approaches concerning the expression of geologic knowledge through UML schemata and ontologies have been around for more than a decade. These approaches have yielded resources that concern specific domains, such as, e.g., lithology. We develop a conceptual model that aims at building a digital encoding of several domains of geologic knowledge, in order to support the interoperability of the sources. We apply the devised terminological base to the classification of the elements of a geologic map of the Italian Western Alps and northern Apennines (Piemonte region). The digitally encoded knowledge base is a merged set of ontologies, called OntoGeonous. The encoding process identifies the objects of the semantic encoding, the geologic units, gathers the relevant information about such objects from authoritative resources, such as GeoSciML (giving priority to the application schemata reported in the INSPIRE Encoding Cookbook), and expresses the statements by means of axioms encoded in the Web Ontology Language (OWL). To support interoperability, OntoGeonous interlinks the general concepts by referring to the upper part level of ontology SWEET (developed by NASA), and imports knowledge that is already encoded in ontological format (e.g., ontology Simple Lithology). Machine-readable knowledge allows for consistency checking and for classification of the geological map data through algorithms of automatic reasoning.

Preprint submitted to Computer and GeoSciences *December 27, 2017*

Keywords: geologic knowledge encoding, geologic unit ontology, geodatabase, geological map, conceptual modeling of geologic knowledge, automatic reasoning

1. Introduction

This paper introduces a novel, semantics–informed geologic mapping process for the production of a synthetic geologic map of a large administrative region, concerning an orogenic system, namely the Geological Map of Piemonte, in the Alps-Apennines interference zone (Piana et al., 2017)[1]. The task of geologic mapping requires the identification of the conceptual objects, or features, with two types of factors that control data-quality:

1. the accuracy of observation/measurement, such as, e.g., the geographic position or the composition of some feature, and
2. the suitability of the representation for the task at hand, such as, e.g., the descriptive elements of some feature.

Here we focus on the latter point, that is the representational issues that raise in the geologic mapping task. In particular, this paper presents a conceptual model that addresses bodies of materials in the Earth, named "geologic units". Geologic units are 1) hierarchically organized into component units, with the most basic units including some compositions of Earth materials and 2) defined according to some basis (which can be chronological, lithological, etc.). The conceptual model provides a data organization: on the one hand, it is compliant with the general knowledge about the geologic units (the objects of the geomapping task); on the other, it contributes to achieve the objective of the task, a classification of the objects with the purpose of their representation on the map (as a graphic object or as a part of an informative system), following an established model of geotectonic evolution of the mapped region. The conceptual model encodes the geologic knowledge to yield a terminological base for the geologic units; the paradigm of linked data (Bizer et al., 2009) supports interoperability of several knowledge sources while keeping the same sources non redundant (see, e.g., the

[1]For a review of the geology of the Alps-Apennines orogenic system, see (Mosca et al., 2009; Beltrando et al., 2010; Dal Piaz, 2010; d'Atri et al., 2016; Molli et al., 2010).

5* deployment schemata for open data[2]); machine-readability of the encoding supports the applicability of automatic reasoning mechanisms, with the goals of consistency checking and instance classification (through Description Logic – DL – formalism (Nardi and Brachman, 2003; Baader et al., 2007) – here expressed in Web Ontology Language OWL 2 (Hitzler et al., 2009a), and reasoning tools – we employ Pellet (Sirin et al., 2007)).

However, the design and implementation of a conceptual model is not straightforward. When semantics comes into play, Earth scientists and computer scientists must address philosophical issues. The principles for data organization raise classical ontological questions such as:

- Are the data at hand instances of general concepts (also called categories or classes)? And how do we motivate the existence of such classes and not others?
- How do we define a correct classification of instances?
- What is the nature of relations existing over classes and instances?

Ontological representation has been the goal of philosophical disciplines for centuries and then of computer science for decades (Hitzler et al., 2009b). The definition and usage of the Semantic Web framework (Berners-Lee et al., 2001) has envisioned a web with a relevant role of the deep meaning of objects, beyond the mere textual format. In particular, a number of languages that are suitable for knowledge representation and reasoning have been developed and tested over several domains. Description logic, implemented through a number of profiles of the Web Ontology Language (OWL) family, interprets the world as classes and instances together with relations (or properties) that provide class restrictions. Such languages are suitable for the classification task that is relevant in geologic mapping and can provide 1) consistency and interoperability of data, 2) a semantic approach to the representation, and, through the machine-readable encoding, 3) an immediate support to applications.

The knowledge sources for realizing such an encoding of classes and instances of the geologic mapping task are 1) the GeoScience Markup Language schemata and vocabularies, 2) the INSPIRE Data Specification on

[2] `http://5stardata.info/en/`

Geology directives, 3) the machine-readable encoding provided for some specific domain, such as the lithology domain (vocabulary Simple Lithology) and the geochronologic time scale (ontology "gts"), and finally 4) for the upper level knowledge, shared across several geologic domains, the upper part of the NASA SWEET ontology. The goal of this paper is to encode the statements reported in a number of authoritative sources into an interlinked machine–readable format; the result is a set of merged ontologies named OntoGeonous[3]. The source statements that are mostly expressed in natural language have been encoded through a process of semantic interpretation that has produced axioms in the OWL–2 language; the concepts and the relations referred to by the axioms are kept coherent in their meaning throughout the whole knowledge base (internal coherence) and with respect to external sources that were already encoded and that are imported into OntoGeonous (external coherence); the geomapping data are classified according to the ontology, consistency checking and novel knowledge inference is achieved through automatic reasoning. We consider our contribution an initial step for the geological knowledge to participate into the Linked Data challenge (the web as one big interlinked database). In large practical applications, our OWL-based approach will likely be replaced by RDF-based syntax and software architecture that scale to data warehouse and continuously changing data (Polleres et al., 2013).

The paper is organized as follows. The next section states the motivations for this work. In section 3, we report on some relevant related work. Section 4 describes the realization of the semantics–informed mapping. Section 5 presents our conclusions. In the following we will use a few schemata. In Figure 1 is the legend of the figures to come.

2. Motivations for this work

In this section, we introduce the data representation of the geologic units of the Piemonte Geological Map (Piana et al., 2017) and how the conceptual modeling can improve such representation. We go through an example

[3]For purposes of proof of concept, the current ontology can be retrieved at the URL: `http://www.di.unito.it/~vincenzo/ontologies/20161013_OntoGeonous_Merge_Inst.owl`, together with a human-readable version of it `http://www.di.unito.it/~vincenzo/ontologies/OntoGeonous.htm`. We will address the issue of url persistence in the near future, after the establishment of an effective general workflow.

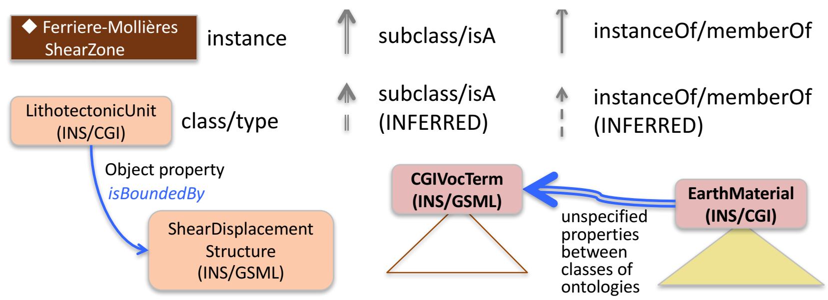


Figure 1: In the figures of this paper: sharp corner boxes with dark background and white text prefixed with diamonds are instances (e.g., Ferriere-Mollires ShearZone); rounded corner boxes with light background and black text are types or classes (e.g., LithotectonicUnit (INS/CGI)); unlabelled solid (dotted in the case of inferences) double vertical arrows are subclass (or isA) relations; unlabelled solid (dotted in the case of inferences) single vertical arrows are instanceOf (or memberOf) relations; curved labelled solid (dotted in the case of inferences) blue arrows indicate that there are Object Property relations between the classes; large curved unlabelled solid double blue arrows indicate that there a number of object properties hold over the classes of two ontologies. Triangles with some root class (e.g., CGIVocTerm) represent ontological encoding of some knowledge source.

from our geologic mapping task and we employ the major knowledge sources mentioned above to produce an item in the underlying data base[4]. The example concerns a specific geologic unit named “Formazione di Baldissero” (Baldissero Formation). If we employ the GeoSciML vocabularies and the INSPIRE directives (see references below), we can list the XML statements in the Listing 1.
“Formazione di Baldissero” is a geologic unit, with an identifier (gml:id, line 03), reported after the namespaces involved (xmlns), a description and a name (both in Italian, original language of the geomapping database, lines 04 and 05), and an occurrence in the map (line 06). It has a geologic history (lines 07–11), here related to one or more geologic events (not furtherly specified). Its type is the lithostratigraphic unit (lines 12–14), whose definition is at a precise URL in the CGI vocabulary of the GeologicUnitType. It is composed (gsmlb:composition) of two parts (gsmlb:CompositionPart), lines 17–31 and lines 34–48 respectively, each with a specific role (stratigraphic

[4]The current encoding is underlying the visualization accessed at the url `http://arpapiemonte.maps.arcgis.com/apps/webappviewer/index.html`

Listing 1 Example of geologic mapping for the geologic unit Formazione di Baldissero, encoded in XML format, with tags from GeoSciML vocabularies.

```
<gsmlb:GeologicUnit
 <!-- all xmlns required -->
 gml:id="Formazione_di_Baldissero">
 <gml:description>Successioni arenaceo-pelitiche e marnose burdigaliano-langhiane.</gml:description>
 <gml:name>Formazione di Baldissero</gml:name>
 <gsmlb:occurrence gml:id="BAD_MF1"/>
 <gsmlb:geologicHistory>
  <gsmlb:GeologicEvent gml:id= ... >
   <!-- geologic event attributes -->
 </gsmlb:GeologicEvent>
</gsmlb:geologicHistory>
<gsmlb:geologicUnitType
 xlink:href="http://resource.geosciml.org/classifier/cgi/geologicunittype/lithostratigraphic_unit"
 xlink:title="lithostratigraphic unit"/>
 <!-- There are two component lithologies in this example -->
 <gsmlb:composition>
  <gsmlb:CompositionPart>
   <gml:name>Formazione di Baldissero CP1</gml:name>
   <gsmlb:role
    xlink:title="stratigraphic_part"
    xlink:href="http://inspire.ec.europa.eu/codelist/CompositionPartRoleValue/stratigraphicPart"/>
   <gsmlb:material>
    <gsmlb:RockMaterial gml:id="Areniti_ibride_Baldissero_RM1">
     <gsmlb:lithology
      xlink:href="http://inspire.ec.europa.eu/codelist/LithologyValue/arenite"
      xlink:title="arenite"/>
    </gsmlb:RockMaterial>
   </gsmlb:material>
   <gsmlb:proportion>
    <!-- what pertains proportions of materials -->
   </gsmlb:proportion>
  </gsmlb:CompositionPart>
 </gsmlb:composition>
 <gsmlb:composition>
  <gsmlb:CompositionPart>
   <gml:name>Formazione di Baldissero CP2</gml:name>
   <gsmlb:role
    xlink:title="stratigraphic_part"
    xlink:href="http://inspire.ec.europa.eu/codelist/CompositionPartRoleValue/stratigraphicPart"/>
   <gsmlb:material>
    <gsmlb:RockMaterial gml:id="Marne_con_intercalazione_arenacee_Baldissero_RM2">
     <gsmlb:lithology
      xlink:href="http://inspire.ec.europa.eu/codelist/LithologyValue/impure_carbonate_sedimentary_rock"
      xlink:title="impure_carbonate_sedimentary_rock"/>
   </gsmlb:RockMaterial>
   </gsmlb:material>
   <gsmlb:proportion>
    <!-- what pertains proportions of materials -->
   </gsmlb:proportion>
  </gsmlb:CompositionPart>
 </gsmlb:composition>
</gsmlb:GeologicUnit>
```

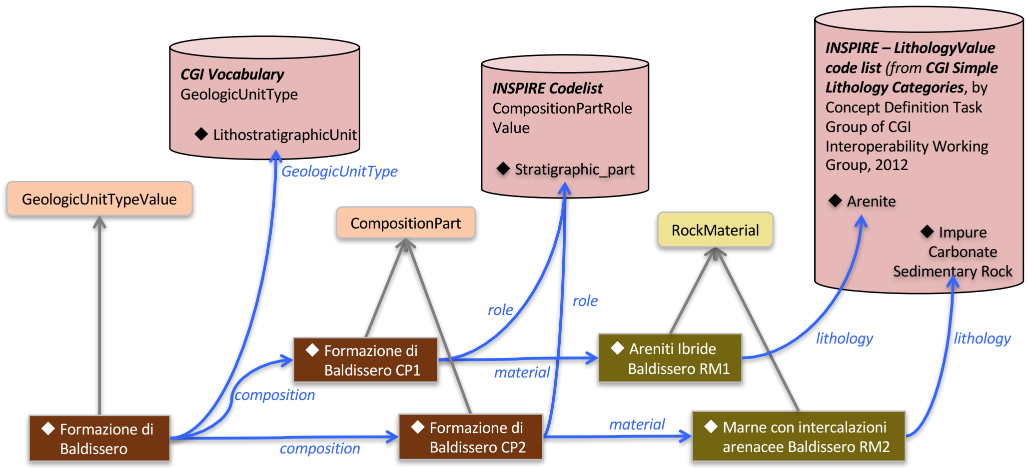


Figure 2: Schematic representation of the GeoSciML encoding for the geologic unit "Formazione di Baldissero" (Baldissero Formation, bottom left corner), with two composition parts, made of materials hybrid arenite and marl with interbedded arenite, respectively.

part in both cases, lines 20 and 37 respectively) and material (some lithology, lines 24 and 41 respectively). Each composition part occupies some proportion of the total (not reported in this example, lines 28–30 and 45–47 respectively). Figure 2 shows a schematic representation of such metadata, in which we have made explicit the connections that are positionally represented in the XML representation over the instances and the types.

The geomapping task requires a framework for the adequate description of the elements in the Listing 1. However, in the XML representation, types or classes (gml tags[5]) have not an explicit definition and the several concepts are not formally interconnected. Values for descriptions should be searched in the mostly informal external resources (CGI vocabularies, INSPIRE codelists, . . .), which are not verified automatically for possible inconsistencies or overlaps. The contribution of this paper is to introduce an interlinked machine–readable encoding of geologic knowledge to serve as a consistent terminological base for the geomapping task. Figure 3 shows a

[5]The OpenGIS Geography Markup Language Encoding Standard (GML) is a XML grammar for expressing geographical features - `http://www.opengeospatial.org/standards/gml`.

schematic representation of the same geologic unit of Figure 2 ("Formazione di Baldissero") in the OntoGeonous encoding. Tags are not mere strings, but references to logical concepts (also called classes) inserted into a large knowledge base. To prevent redundancy, classes are organized hierarchically through the principle of set inclusion (or isA relation, represented by the triangles). Whenever possible from the authoritative sources, we introduced class definitions, which state the necessary and sufficient conditions for the class existence and are paramount for the automatic classification task over instances. Classes belonging to external specific ontologies are not re-encoded; though according to the linked data paradigm we can refer to such classes from the OntoGeonous ontology through some IRI (Internationalized Resource Identifier), in the current implementation, we directly imported the whole external ontology for prototype validation. The several sources mentioned above, which were referred through URL's to specific concepts, are now interconnected and reasoning mechanisms can be applied to check the knowledge consistency at large and to classify instances according to the relations that hold over instances. This encoding of community standards as well as of the instances in the map is a step towards interoperability: another geomapping process would refer to the same knowledge base, favoring consistency of representations and comparisons over several projects, with mutual benefits in terms of ease of geomapping implementation and of application/services development.

3. Related work

The sources that make up the backbone of our approach are addressed later in the paper. Here, we refer to a number of approaches that apply a semantics–informed interpretation of datasets (especially in the context of geomapping tasks) and that we have taken into account during our research. We address three types of related works: the technical infrastructures for semantics–informed applications, the ontological encoding of specialized domains, and the usage of authoritative resources (such as GeoSciML and INSPIRE).

The technical infrastructures are very numerous in the geomatic literature. They are complementary to OntoGeonous: where they introduce technicality for realizing services, we introduce content (or knowledge) to support those services. Eventually, in general, all these infrastructures could benefit

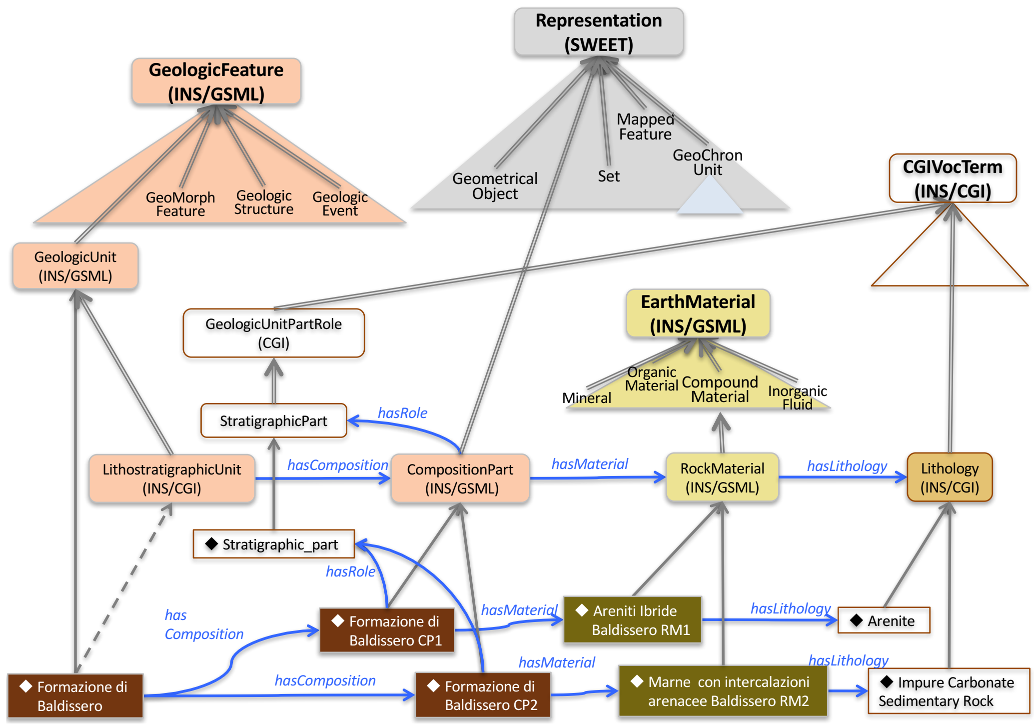


Figure 3: OntoGeonous encoding of the geologic unit of Figure 2 in a schematic representation.

from the inclusion of OntoGeonous as an authoritative knowledge base. Here we mention just a few, related to semantics–informed applications.

Geon[6] is an open collaborative project that develops a cyber–infrastructure for the integration of 3D– and 4D– data, where formal ontologies (SWEET, among others) are used to coordinate and integrate conceptual schemas of heterogeneous geological maps (cf. (Ma, 2011)). Project Geon developed the OpenEarth Framework, a semantics–based toolsuite for integration and visualization of multi-dimensional data (Ludäscher et al., 2003, 2008).

GeoBrain[7] is a multidisciplinary system aimed at popularizing NASA data and information through knowledge management technologies, covering spatiotemporal factors, physical facts, disciplines and platforms, in reference

[6] `http://www.geongrid.org/` and its evolution `http://www.opentopography.org/`
[7] `http://geobrain.laits.gmu.edu/`

to ontology SWEET (Zhao et al., 2009). OntoGeonous could be a domain ontology in this application.

AuScope[8] is an integrated national framework that uses vocabulary–based services for querying geological maps (Woodcock et al., 2010). The British Geological Survey (BGS) has developed and implemented a cyber-infrastructure that makes explicit much of the implicit knowledge acquired by new geological surveys (Howard et al., 2009). SETI (Semantics Enabled Thematic data Integration)(Durbha et al., 2009) is a system that enables the retrieval of information from thematic data archives via semantics–driven searches. In these projects, ontologies were developed for the classification schemes and a shared-ontology approach for integrating the application level ontologies; however, they are not available for further usages and consistency checking has not been an issue in these projects.

More restricted in focus are CHRONOS (Fils et al., 2009), which integrates stratigraphic databases, and Hydroseek (Beran and Piasecki, 2009), an ontology–aided search engine, that allows users to query multiple hydrologic repositories, with a knowledge base that covers water quality, meteorology and hydrology domains.

Finally, related to Ma's ontology mentioned above is the pilot interactive multimedia project developed by (Ma et al., 2012), who provided an animated visualization and interaction functions over the Geologic Time Scale ontology (Ma, 2011). OntoGeonous could be used for connecting specific knowledge with general geologic knowledge; however, this would require an adaptation of the present ontologies for the sake of the interoperability goal.

Approaches aimed at the ontological encoding of specialized domains are Virtual Solar–Terrestrial Observatory (VSTO) and Space Physics Archive Search and Extract (SPASE). VSTO[9] is a semantic data framework based on an ontology of the domains of solar physics, space physics and solar-terrestrial physics (Fox et al., 2009). As in the case of OntoGeonous, VSTO also refers to the functional decomposition of SWEET, reusing, e.g., the notions of Earth and sun realms, respectively. The SPASE consortium[10] have been creating a comprehensive space physics data model (Narock et al., 2009), converted into an OWL ontology, consists of agreed–upon terminology

[8] https://www.researchgate.net/publication/234183449_AuScope's_use_of_Standards_to_Deliver_Earth_Resource_Data

[9] https://www.vsto.org

[10] http://www.spase-group.org/

and –definitions for use in the community and use in virtual observatories.

These approaches employ ontological encoding of specialized domains; as such, these ontologies approach the terminological problem within some separate domain, with limited inter-connections or integrated applications. OntoGeonous could embed the data model here built to provide interconnections upon all the branches of geologic knowledge, improving consistency and interoperability.

Finally, there are a number of approaches that make the effort of relying on authoritative resources (such as GeoSciML), without introducing ad hoc knowledge specifications. All these approaches currently make a very basic use of ontological encoding: OntoGeonous improves such methods by providing a comprehensive approach to the formal encoding of the geologic knowledge, aimed at subsequent automatization of application algorithms. OneGeology[11] has the goal of creating a worldwide geological map by harmonizing data from different providers, using GeoSciML standard. TaxonConcept[12] (Huber and Klump, 2009) allows to store Open Nomenclature synonymy lists (list of citations related to a taxon name), in the field of taxonomic classification of fossil species. The United States Geoscience Information Network[13] aims to facilitate the access to geoscience information provided by state and federal geological surveys of the United States, with GeoSciML as data transfer standard (Richard and Allison, 2016).

The approach described in this paper departs from such initiatives in contributing to an integration of the knowledge sources in the terms of a machine–readable encoding, addressing the convergence on a shared knowledge kernel. In order to make things concrete, the encoding is immediately applied to the geomapping task to demonstrate the usefulness and the feasibility of the enterprise.

4. Realization of OntoGeonous

OntoGeonous is a merged ontology consisting of a number of ontologies, some realized anew and some already existing: this implements the paradigm

[11] `http://portal.onegeology.org/OnegeologyGlobal/` and `http://onegeology-europe.brgm.fr/geoportal/viewer.jsp`

[12] `http://taxonconcept.stratigraphy.net/`

[13] `http://www.dgs.udel.edu/projects/united-states-geoscience-information-network-usgin` and `http://usgin.org/`

of linked data and avoids the re-encoding of existing machine–readable knowledge.

The knowledge sources we have taken into account are the statements, schemata, vocabularies, and encoded ontologies, from major authoritative institutions (Table 1 summarizes the markers that identify the sources):

- GeoScience Markup Language (GeoSciML)[14] expressed in a number of UML schemata (classes, features, attributes, associations) and statements in natural language, to be encoded in OWL;

- INSPIRE (Infrastructure for Spatial Information in the European Community)[15] aimed at creating a European Union spatial data infrastructure, expressed through natural language statements, to be encoded in OWL;

- SWEET (Semantic Web for Earth and Environmental Terminology)[16], developed by NASA–Jet Propulsion Laboratory since 2002, a set of ontologies for environmental and Earth system science terms (Raskin and Pan, 2005; Barahmand et al., 2010), expressed in OWL;

- vocabularies of specific subdomains of geologic knowledge that are relevant for the geomapping task[17], encoded in the SKOS format (Simple Knowledge Organization System[18]) and available in .rdf and .ttl versions. For example, we have imported the lithology domain vocabulary named Simple Lithology[19], through a simple encoding that creates taxonomic classes as translated from narrower/broader relations over individuals. For the geological timescale, we have integrated ICS Geological Time Scale Ontology (Ma, 2011) as a subtaxonomy of the Geochronologic Unit class of SWEET Representation. In particular, the Geochronologic Unit class of OntoGeonous corresponds to

[14] Version 4.0 (2015), http://www.geosciml.org

[15] D2.8.II.4 INSPIRE Data Specification on Geology Technical Guidelines v. 3.0. (10.12.2013) (http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_GE_v3.0.pdf)

[16] (https://sweet.jpl.nasa.gov/

[17] http://resource.geosciml.org/vocabulary/cgi/201211/

[18] https://www.w3.org/2004/02/skos/

[19] http://resource.geosciml.org/vocabulary/cgi/201211/simplelithology.rdf

SWEET GeologicTimeUnit class (actually the hierarchical path Representation – NumericalEntity – Interval – Duration – GeologicTimeUnit). We selected Ma's ICS Geological Time Scale because, in spite of the simplicity of encoding, it allows the inheritance of a large number of attributes (multilingual thesaurus, ICS standard RGB code, relations between concepts). For a more complete ontological approach, we are considering to integrate Cox and Richard's GTS ontology in the future (Cox and Richard, 2015).

Authoritative source	**Annotation string**
GeoSciML schemata	"GSML"
CGI vocabularies	"CGI"
INSPIRE	"INS"
CGI and INSPIRE shared	"CGI-INS"
GSML and INSPIRE shared	"GSML-INS"
International Commission on Stratigraphy	"ICS"

Table 1: Suffixes for concept terms to mark the provenance from some authoritative source.

Once we have identified the domain elements that are relevant for the geomapping task, the steps for the realization of OntoGeonous have been the following:

1. taxonomization, that is the identification of the subsumption relation over classes inferred to exist from the general schemata and vocabularies;
2. concept axiomatization, that is the introduction of definitions of concepts, i.e. statements that define a concept through the enumeration of necessary and sufficient conditions for its existence; the goal here is the issue of disambiguation within the classification task, that is the possibility of unambiguously classifying some object; when this is possible, we are able to implement automatic reasoning and then classification;
3. incremental validation of knowledge through the encoding of examples drawn from the map and automatic verification of consistency with respect to the whole knowledge base.

In our case, the objects that result from the conceptual modeling task are the geologic units, accurately identified on the map, bordered by geologic structures and related to geologic events. In the following, we address the individual encoding phases as separate, linearly ordered processes. However, the real encoding has proceeded through several adjustments in parallel on the several phases.

4.1. Identification of knowledge sources and big picture

Figure 4 illustrates a schematic interconnection of the knowledge sources that compose OntoGeonous. The triangles represent the major concept taxonomies, concerning different realm (kept distinct by colors). In the upper left corner, the original sources: GeoSciML–INSPIRE and SWEET ontology on the left, ICS GTS and Simple Lithology ontologies on the right (notice that the latter two are already in ontological format, OWL file format). The most relevant taxonomy of concepts is provided by GeoSciML–INSPIRE source.

The core of the geologic knowledge is the (orange–colored) taxonomy rooted by Geologic Feature, with four major subclasses, GeoMorphologicFeature, GeologicUnit, GeologicStructure, and GeologicEvent (see below). This taxonomy is connected to all those features, attribute, properties, that constitute generic knowledge, shared with other scientific disciplines. These connections are illustrated as curved blue lines. All the knowledge sources that merge into OntoGeonous make a reference to the frameworks (such as SWEET) that encode the concepts that are abstractions of the specific ones employed in the Earth sciences.

The concept GeologicFeature, which encompasses all the geologic core knowledge, is related to many external concepts, which define its major distinctive attributes. We enumerate these external concepts going downwards on the blue arrows from GeologicFeature in Figure 4. First, GeologicFeature is related to some MappedFeature, a fundamental relation for the geomapping task. A mapped feature is the spatial extent of the geologic feature on the map. In turn, a mapped feature is related to some geometrical object (such as, e.g., a polygon), a subconcept of the generic concept of Representation, in the upper part of the ontology SWEET. Second, GeologicFeature is related to some GeoChronologicUnit, root of the ICS GTS taxonomy (the light blue triangle in Figure 4 – upper right) and identified with the corresponding concept in the Representation taxonomy of ontology SWEET. Finally, GeologicFeature is related to the CGIVocabularyTerm vocabularies (a taxonomy), which provide specific concepts for the several subdomains, such as the ones for the Earth materials, and to the abstract descriptions in GeoSciML, which encode attributes, such as the unit thickness.

GeologicFeature is subdivided into four sub–taxonomies, namely GeoMorphologicFeature, GeologicUnit, GeologicStructure, GeologicEvent. Each of these concepts addresses some distinctive object of the geologic knowledge:

1. GeoMorphologicFeature describes the landforms, which have event processes as their major distinctive attribute. Event processes, which concern the creation, modeling, etc. of geomorphologic features, are described by a taxonomy/ontology whose major subclasses are NaturalEarthProcess and HumanActivity. The event process taxonomy can be considered as a mid–level ontology subsumed by the concept Process (in turn, subclass of Phenomenon) in the SWEET ontology.
2. GeologicUnit describes a body of some material, which has the composition material as distinctive attribute. As it happens with EventProcess, also EarthMaterial, which specifies the Substance concept in the SWEET ontology and includes the ontology SimpleLithology, is a taxonomy with a number of subclasses and related vocabularies (CGIVocabularyTerm taxonomy and GSML Abstract Description). In particular, CompoundMaterial, a subclass of EarthMaterial, is the object of CompositionPart, an intermediate representation concept that addresses the splitting of some body of material into several parts according to their composition materials.
3. GeologicStructure describes the configurations or patterns in which the geologic units are arranged, either internally or externally. In particular, GeologicStructure is mainly described through some abstraction, such as inhomogeneity, internal deformation, pattern, or some actual features such as fracture or fault, occurring in the Earth material.
4. GeologicEvent describes the relevant events in geology. Given the INSPIRE definition as "an identifiable event during which one or more geological processes act to modify geological entities" and that "should have a specified geologic age and process, and may have a specified environment", we assume that a GeologicEvent is characterized by both an EventProcess and an EventEnvironment. The latter two are subclasses of the PlanetaryRealm and Phenomena concepts in SWEET, respectively, and refer to specific vocabularies in GeoSciML.

4.2. Taxonomization of concepts and criteria of subsumption

Each of the four major concepts is then developed into a taxonomy. In this section, we illustrate the taxonomy of the Geologic Unit (see Figure 5) by

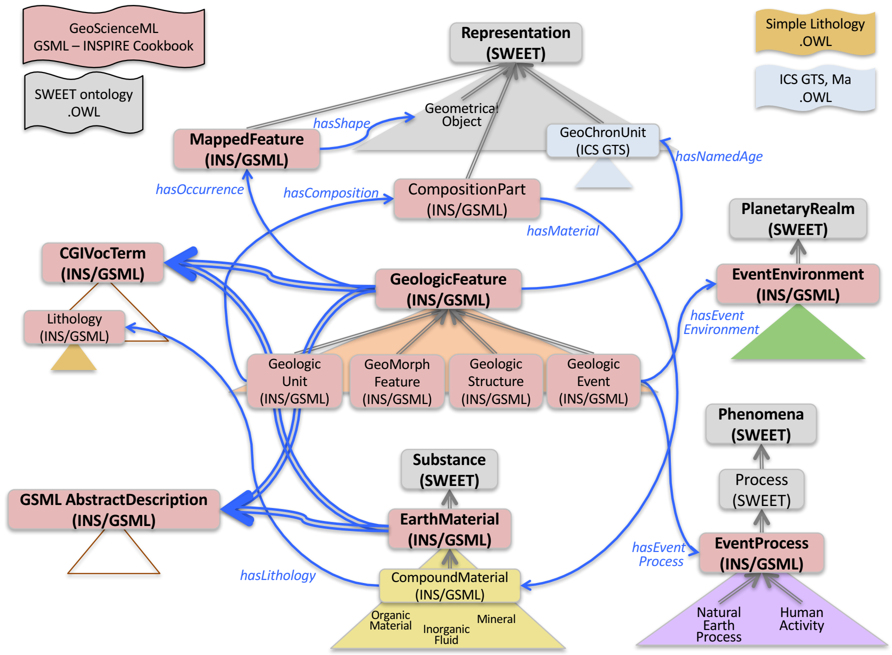


Figure 4: The interlinked geologic knowledge base OntoGeonous at a glance: main isA relations (double-line arrows, e.g., GeologicUnit isA GeologicFeature), object property relations over classes (e.g., a GeologicFeature hasOccurrence some MappedFeature), unspecified object properties between the classes in two taxonomies (e.g., classes within the taxonomy rooted by GeologicFeature and within the taxonomy rooted by GSML-AbstractDescription). Colors distinguish the provenance of the classes from the individual authoritative resources.

addressing the criteria for defining the subclass, or subsumption, relation. In proceeding from classes to subclasses, it is useful to refer to some parameter that can provide some form of partition over the subclasses with respect to the mother class. Although we can have subclasses with more than one parent class, it is helpful to provide some criteria for mutual exclusion of subclasses when possible, to prevent ambiguity in inheritance procedures: this makes the classification mechanism more effective, with advantages onto the geomapping task.

The taxonomy of the geologic units in Figure 5 has been encoded from the CGI/INSPIRE sources. The schema illustrates the major factors that keep the several subclasses distinct, as they are introduced by the linguistic expression “is defined on the basis of”, which recurs regularly in CGI/INSPIRE definitions. This happens because, though a geologic unit can in principle belong to several classes, there are preferred factors that determine its actual classification. For example, a unit can be bounded by a shear displacement structure as well as contain fossils; so, it can be classified preferably on the basis of either the type of its bounding geologic structure or the type of its fossil content; the geologist usually takes such decision according to her/his classification task and the knowledge encoding must support such decision. An interesting future research area could be the devise of heuristics for establishing such preferences: now the system reasons on whatever property has been encoded for some instance and generally yields multiple classifications for it.

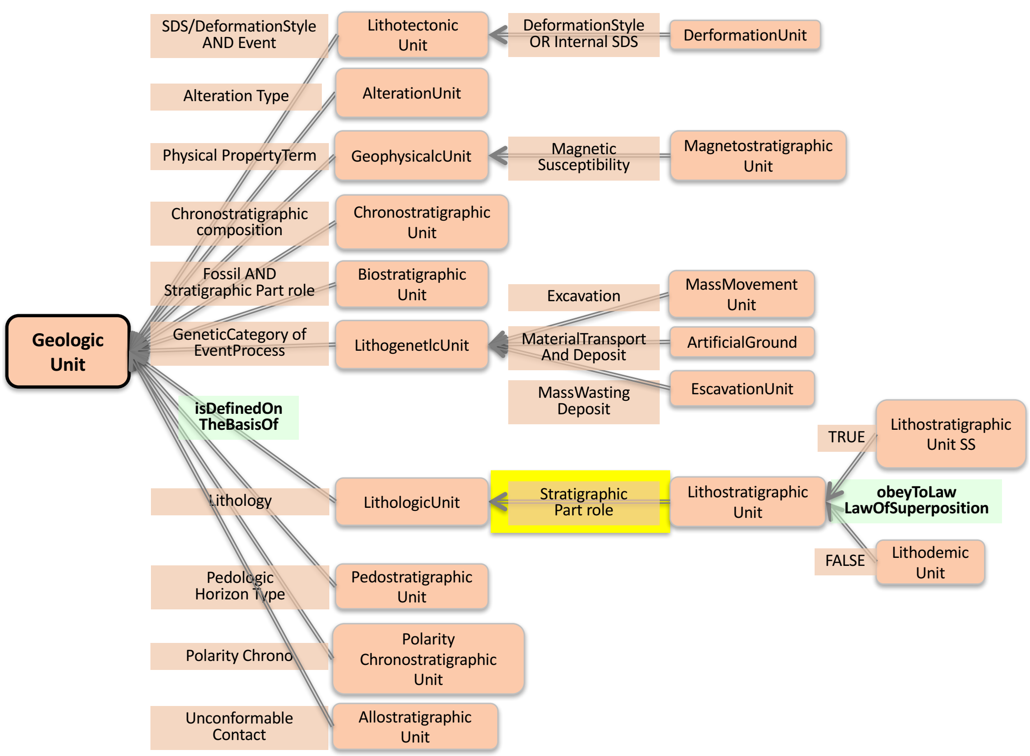


Figure 5: The criteria for subclasses of geologic unit.

4.3. Concept axiomatization of major classes

The concept axiomatization process is a fundamental part of the ontological encoding because of its relevance for the classification task. The goal of this process is to produce an axiom, that is an absolute truth about a concept: operationally, this means to identify the necessary and sufficient conditions for an object to be classified as an instance of some concept. This is why a concept is often called a class in the modern ontological terminology. In order to illustrate the concept axiomatization process, which goes through semi–formal steps of semantic interpretation of natural language definitions and UML schemata, we introduce a running example (Lithotectonic Unit).

First, we select the relevant statements from the knowledge sources. For the example of the Lithotectonic unit, the main knowledge sources are the INSPIRE directive (GeologicUnitTypeValue[20]) and the CGI GeologicUnitType vocabulary[21]. The definition reported in INSPIRE is:

> Geologic unit defined on basis of structural or deformation features, mutual relations, origin or historical evolution. Contained material may be igneous, sedimentary, or metamorphic.

Second, on the basis of such statement, possibly merged with expressions from other knowledge sources, we produce a *protoaxiom*. A protoaxiom is a statement expressed in a controlled natural language: the table 2 reports schematically the protoaxiom production process for the case of the Lithotectonic unit.

The fact that a Lithotectonic unit is a Geologic unit of some sort is translated into the fact that a Lithotectonic unit is a subclass of the Geologic unit class (table header). The notion of equivalence (EQUIVALENT TO) corresponds to the notion of definition, that is in providing the necessary and sufficient conditions for classification. The conditions are in the third and fourth rows of the table, where we can find, on the left (the first column), the expressions in natural language and, on the right (the second column) the expression in pseudo–logic language, that make use of restrictions (object properties – OP and datatype properties – DP) over classes.

In the third row, the expression

[20]`http://inspire.ec.europa.eu/codelist/GeologicUnitTypeValue/lithotectonicUnit/`

[21]`http://resource.geosciml.org/classifier/cgi/geologicunittype/lithotectonic_unit`

INS - CGI: "Geologic unit"	subclass of CLASS GeologicUnit- GSML/INS
	EQUIVALENT TO
INS - CGI: "defined on basis of structural or deformation features, mutual relations, origin or historical evolution"	"Structural features": OP isBoundedBy some class ShearDisplacementStructure GSML/INS OR "Deformation features": OP hasDeformationStyle some class DeformationStyle - CGI OR "Origin or Historical evolution": OP isRelatedToEvent some class GeologicEvent–GSML/INS NOTE: "Mutual Relations" interpreted as spatial relations imposed by a SDS, i.e. OP isBoundedBy class ShearDisplacementStructure)
INS - CGI: "Contained material may be igneous, sedimentary, or metamorphic"	hasComposition some class CompositionPart - GSML AND hasMaterial some class CompoundMaterial - GSML (inherited from CLASS GeologicUnit - GSML/INS) NOTE: igneous + sedimentary + metamorphic = class CompoundMaterial (IGNORED)

Table 2: Construction of the protoaxioms: left column: expression from the information source; right column: protoaxiom expressed in pseudo–Manchester syntax style.

> ... defined on basis of structural or deformation features, mutual relations, origin or historical evolution.

is split into several parts that are intended as the conjunctive terms of the definition: "structural or deformation features", "mutual relations", "origin or historical evolution". The first part is in turn subdivided into "structural features" and "deformation features", intended as possible alternatives (not necessarily exclusive). "Structural features" can be interpreted as "a geologic unit that is bounded by a shear displacement structure": this is encoded as a restriction on the GeologicUnit class through the object property `isBoundedBy`, whose range is the GeologicStructure subclass `ShearDisplacementStructure`. Similarly, "deformation features" can be interpreted as "a geologic unit that has some form of deformation style": this is encoded again as a restriction on the GeologicUnit class through the object property `hasDeformationStyle`, whose range is the vocabulary derived class `DeformationStyle`. The second part, "mutual relations" is included in the "structural features" interpretation as "the spatial relations imposed by the related geologic structure", and so does not contribute further to the definition. Finally, the third part, "origin or historical evolution", can be interpreted as a generic relation to some geologic event, through the object property `isRelatedToEvent`, whose range is the generic class `GeologicEvent`.

The fourth row makes reference to the composition material of the geologic unit. Though the right column reports an encoding in terms of class restrictions, as reported in the note, we interpreted the statement as redundant, since it reports all the possible materials, and decided not to add any logic statement to the previous definition.

Third, the protoaxiom is encoded in OWL language, to form the axiom. The example of axiom concerning the Lithotectonic Unit is the following:

```
CLASS LithotectonicUnit  CGI/INS    EQUIVALENT TO
CLASS GeologicUnit - GSML/INS and
((hasDeformationStyle some DeformationStyle) or
 (isBoundedBy some ShearDisplacementStructure))
and
(isRelatedToEvent some GeologicEvent)
```

Notice that the connectives `and/or` are nested in the representation above: in fact, deformation style and shear displacement structure can be alternative (though also co–existent, inclusive `or`), while the relationship with some event is necessary for the definition. Figure 6 shows a graphic representation of the axiom.

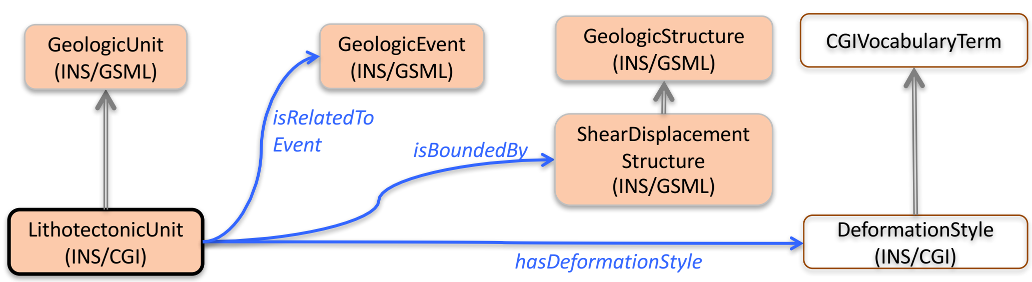


Figure 6: Axiom of the lithotectonic unit in graphic format. The defined class is in bold; the reported object properties are the ones that define the class.

4.4. Encoding of instances and incremental validation of knowledge

Each time a novel axiom is added to the knowledge base, some instances that are related to the axiom are encoded to test the consistency through an application of automatic reasoning. In Figure 7 we report the encoding

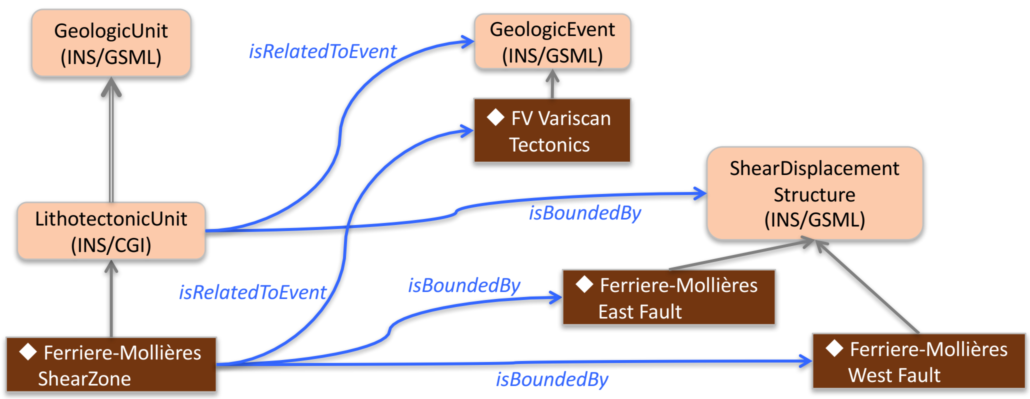


Figure 7: Encoding an instance of geologic unit from the map. The identifiers prefixed with a diamond, in white text on dark background, are instances of the classes connected to them through upward–directed simple arrows.

of one instance of Lithotectonic unit, namely the Ferriere–Mollières Shear Zone, which is bounded by two faults and is related to a tectonic event.

The consistency of the knowledge base is tested through the application of automatic reasoning techniques, which reveal possible inconsistencies and infer novel knowledge. Figure 8 shows two inferences employed to verify the consistency of the knowledge base. Ferriere–Mollières Shear Zone is created as instance of the generic class GeologicUnit and engaging into object properties of isBoundedBy, hasDeformationStyle, and isRelatedToEvent types, respectively. According to the definition above, such an instance is classified automatically as a Lithotectonic unit and, in turn, as a Deformation unit, because it is both inferred as Lithotectonic and restricted by the "hasDeformationStyle" property (cf. taxonomy in Figure 5). This result shows that the reasoning mechanism can support the filling of the database and check the consistency of the knowledge base as it grows, incrementally.

Currently, the OntoGeonous ontology contains 707 concepts, split into the core ontology of the geologic features (and geologic units in particular, while still lacking geologic structures, geomorphologic features, geologic events), the Earth materials, the geochronologic units, the environments and the events, the upper level concepts equalled to SWEET upper concepts (cf. the big picture in Figure 4). Concepts are restricted through 100 object properties, which connect some concept to some other concept, mainly em-

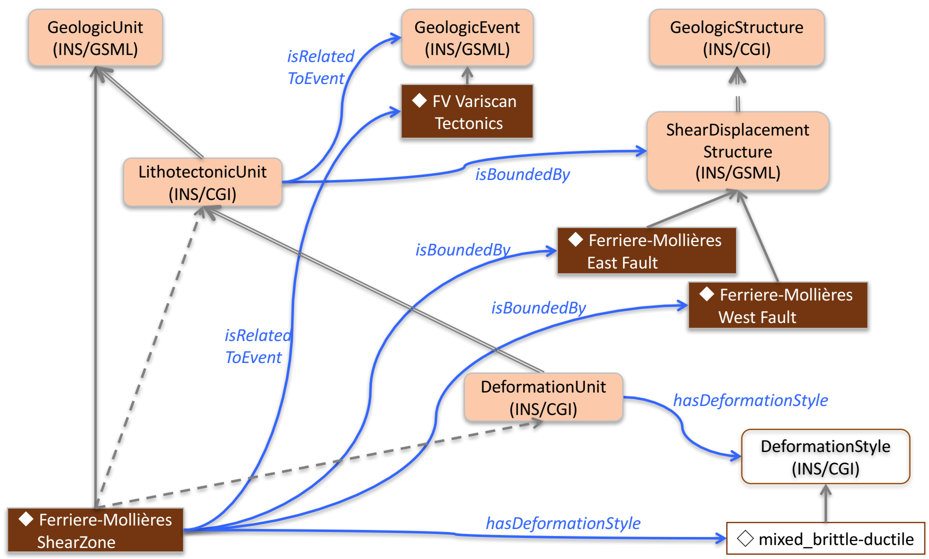


Figure 8: Encoding of the example (solid arrows) and automatic classification (hyphenated arrows marked in yellow).

ployed for axiom definition (a geologic unit is a geologic feature restricted to have some composition of bodies), and 41 datatype properties, which connect some concept to some attribute (e.g., a boolean value – true/false – representing that the law of superposition holds). We have introduced 83 equivalence axioms, that is concept definitions that state the necessary and sufficient conditions for the existence of some class.

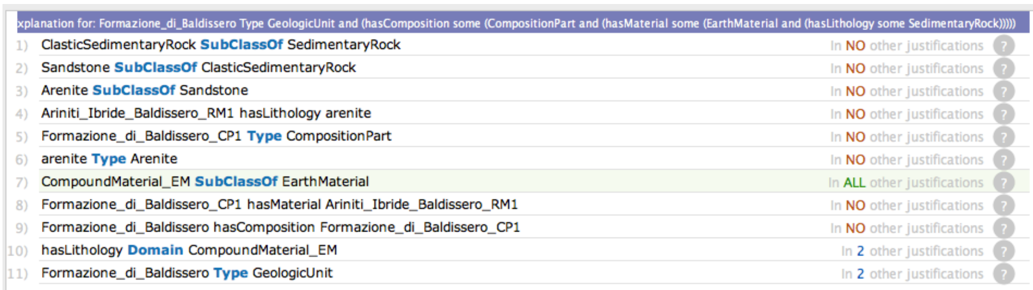
In order to classify the instances of geologic units in the Piemonte geological map, with their Earth materials, the geochronologic unit associated, the geologic structures that bound the units, the geologic events that originated the units, we have currently introduced 520 instances. Of such instances, 34 are geologic units (over a totality of about 6,000 geologic units in the map). These 34 units were selected to cover the most of the classes contained in the ontology; the rest of the instances account for all the concepts that contribute to the definitions of the unit classes. We encode the rest of the units through an ingestion program that creates the instances after a direct retrieval from the current data base underlying the map.

We conclude this section with one example of query on the current knowledge base. If we pose OntoGeonous the query "get all the instances that are

GeologicUnit and have a sedimentary rock composition", that is encoded as

```
GeologicUnit and
(hasComposition some (CompositionPart and
  (hasMaterial some (EarthMaterial and
    (hasLithology some SedimentaryRock)))))
```

we get as result the instance "Formazione di Baldissero". The Figure 9 reports the explanation for the result: the instance with the identifier `Formazione_di_Baldissero` (Baldissero Formation) is a geologic unit (row 11), that has the composition part instance `Formazione_di_Baldissero_CP1` (rows 9 and 5), whose material is `Areniti_Ibride_Baldissero_RM1` (Baldissero Hybrid Arenite, row 8); `Areniti_Ibride_Baldissero_RM1` has a lithology instance `arenite` (row 4), whose class is `Arenite`, subclass of `Sandstone`, subclass of `ClasticSedimentaryRock`, subclass of `SedimentaryRock` (rows 3, 2, and 1).


xplanation for: Formazione_di_Baldissero Type GeologicUnit and (hasComposition some (CompositionPart and (hasMaterial some (EarthMaterial and (hasLithology some SedimentaryRock))))

1)	ClasticSedimentaryRock **SubClassOf** SedimentaryRock	In NO other justifications
2)	Sandstone **SubClassOf** ClasticSedimentaryRock	In NO other justifications
3)	Arenite **SubClassOf** Sandstone	In NO other justifications
4)	Ariniti_Ibride_Baldissero_RM1 hasLithology arenite	In NO other justifications
5)	Formazione_di_Baldissero_CP1 **Type** CompositionPart	In NO other justifications
6)	arenite **Type** Arenite	In NO other justifications
7)	CompoundMaterial_EM **SubClassOf** EarthMaterial	In ALL other justifications
8)	Formazione_di_Baldissero_CP1 hasMaterial Ariniti_Ibride_Baldissero_RM1	In NO other justifications
9)	Formazione_di_Baldissero hasComposition Formazione_di_Baldissero_CP1	In NO other justifications
10)	hasLithology **Domain** CompoundMaterial_EM	In 2 other justifications
11)	Formazione_di_Baldissero **Type** GeologicUnit	In 2 other justifications

Figure 9: Explanations for the results of the query "Get all the instances that are GeologicUnit and have a sedimentary rock composition". Screenshot from the Protègè editor.

In this example, we only got one result because of the limited number of instances that currently populate the knowledge base. We are going to fill the knowledge base with several thousands of geological features of the Piemonte Geological Map, in order to offer web services based on the reasoning capabilities we have exhibited here[22].

[22]This service will be hosted on Arpa Piemonte Environmental Agency geoportal - `http://arpapiemonte.maps.arcgis.com/apps/webappviewer/index.html?id=fff173266afa4f6fa206be53a77f6321`)

5. Conclusion

This paper has introduced a deep semantic representation into the geologic mapping process. We have developed a logical encoding of the general geologic knowledge, the OntoGeonous initiative, based on authoritative resources, such as GeoSciML and INSPIRE, and referring to widely accepted upper level ontological concepts (such as the ones reported in NASA SWEET ontology), also importing knowledge that is already encoded in the OWL format (such as Simple Lithology). So, OntoGeonous is a merged set of computational ontologies. The knowledge base has then been applied to the classification of the elements of a geologic map after the development of a suitable conceptual model. Machine–readable knowledge allows for consistency checking, interoperability, and classification of the geomapping data through the algorithms of automatic reasoning.

OntoGeonous has been the product of the interaction between geologists and computer scientists, who exchanged many ideas during the encoding process. During the ontology development, an effective tool for discussion of the axiomatic encoding ongoing was the implementation of a wiki[23]. Now, the wiki is released as a resource for further investigation as well as a human readable version of the knowledge (cf. (Howard et al., 2009) on the importance of wiki's for knowledge creation).

The formal encoding of the geological knowledge opens new perspectives for the analysis and representation of the geological systems. These often have a very complex internal setting and a large range of physical properties, acquired in distinct geochronological steps (punctuated by geologic events), but rarely fully explicitly described (Balestro and Piana, 2007) (Loudon, 2000) (Frodeman, 1995) (Brodaric et al., 2004). In fact, once that the major concepts employed in the implementation of a geological map data base are defined, with their meaning explicitly expressed through a computational ontology, the resulting formal conceptual model of the geologic system can hold across different technical and scientific communities.

6. Acknowledgements

We, the authors acknowledge a grant (direct beneficiary Dario Mimmo) for the Lagrange Project - CRT Foundation / ISI Foundation. A number of

[23] https://www.di.unito.it/wikigeo/

people read and commented draft versions of this paper: Rossana Damiano, Giandomenico Fubelli, Marco Giardino, Daniele P. Radicioni. Sergio Tallone and Luca Barale gave us fruitful suggestions and support. Claudio Mattutino and the Technical Staff of the Computer Science Department gave us support in the setup of the wiki. The four anonymous reviewers gave us very useful suggestions to improve the paper. We thank all of them.

7. References

Baader, F., Horrocks, I., Sattler, U., 2007. Chapter 3: Description logics. In: van Harmelen, F., Lifschitz, V., Porter, B. (Eds.), Handbook of Knowledge Representation. Elsevier.

Balestro, G., Piana, F., 2007. The representation of geological knowledge and uncertainty in databases of GIS geological maps. Boll. Soc. Geol. It. (Ital. J. Geosci.) 126 (3), 487–495.

Barahmand, S., Taheyrian, M., Al-Ashri, S., Upadhyay, 2010. A survey on sweet ontologies and their applications (http://www-scf.usc.edu/ taheriya/reports/csci586-report.pdf). Tech. rep., University of Southern California.

Beltrando, M., Compagnoni, R., Lombardo, B., http://doi.org/10.1016/j.gr.2010.01.009 2010. (ultra-) high-pressure metamorphism and orogenesis: An alpine perspective. Gondwana Research, 1–20.

Beran, B., Piasecki, M., 2009. Engineering new paths to water data. Computers and Geosciences 35, 753–760.

Berners-Lee, T., Hendler, J., Lassila, O., 2001. The semantic web. Scientific American Magazine.

Bizer, C., Heath, T., Berners-Lee, T., 2009. Linked data - the story so far. International Journal on Semantic Web and Information Systems 5 (3), 1—22.

Brodaric, B., Gahegan, M., Harrap, R., 2004. The art and science of mapping: computing geological categories from field data. Computer and Geosciences 30 (7), 719–740.

Cox, S. J. D., Richard, S. M., March 2015. A geologic timescale ontology and service. Earth Science Informatics 8 (1), 5–19.

Dal Piaz, G. V., http://doi.org/10.3809/jvirtex.2010.00234 2010. The Italian Alps: a journey across two centuries of alpine geology. Journal of the Virtual Explorer 36.

d'Atri, A., Piana, F., Barale, L., Bertok, C., Martire, L., issn=1437-3262 2016. Geological setting of the southern termination of western alps. International Journal of Earth Sciences, 1–28.

Durbha, S., King, R., Shah, V., Younan, N., 2009. A framework for semantic reconciliation of disparate earth observation thematic data. Computers and Geosciences 35, 761–773.

Fils, D., Cervato, C., Reed, J., Diver, P., Tang, X., Bohling, G., Greer, D., 2009. Chronos architecture: Experiences with an open-source services-oriented architecture for geoinformatics. Computers and Geosciences 35, 774–782.

Fox, P., McGuinness, D. L., Cinquini, L., West, P., Garcia, J., Benedict, J. L., Middleton, D., 2009. Ontology-supported scientific data frameworks: The virtual solar-terrestrial observatory experience. Computers and Geosciences 35, 724–738.

Frodeman, R., 1995. Geological reasoning: geology as an interpretive and historical science. Geological Society of America Bulletin 107 (8), 960–968.

Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., Rudolph, S., 2009a. Owl 2 web ontology language primer. Tech. rep., World Wide Wed Consortium (W3C).

Hitzler, P., Krötzsch, M., Rudolph, S., 2009b. Foundations of Semantic Web Technologies. Chapman & Hall/CRC.

Howard, A. S., Hatton, B., Reitsma, F., Lawrie, K. I., 2009. Developing a geoscience knowledge framework for a national geological survey organisation. Computers and Geosciences 35, 820–835.

Huber, R., Klump, J., 2009. Charting taxonomic knowledge through ontologies and ranking algorithms. Computers and Geosciences 35, 862–868.

Loudon, T. V., 2000. Geoscience after it: Part j human requirements that shape the evolving geoscience information system. Computers and Geosciences 26 (3A), A87–A97.

Ludäscher, B., Lin, K., Brodaric, B., Baru, C., 2003. Geon: toward a cyberinfrastructure for the geosciences—a prototype for geological map interoperability via domain ontologies. In: Soller, D. (Ed.), Digital Mapping Techniques '03—Workshop Proceedings. Millersville, PA, USA,, pp. 223–229.

Ludäscher, B., Lin, K., Jaeger-Frank, E., Bowers, S., Altintas, I., Baru, C., 2008. Enhancing the geon cyberinfrastructure: Semantic registration, discovery, mediation, and workflows. Tech. rep., Internal Report of the Geosciences Network.

Ma, X., 2011. Ontology spectrum for geological data interoperability. Ph.D. thesis, University of Twente.

Ma, X., Carranza, E., Wu, C., der Meer, F. V., 2012. Ontology-aided annotation, visualization and generalization of geological time scale information from online geological map services. Computers and Geosciences 40, 107–119.

Molli, G., Crispini, L., Malusà, M. G., Mosca, P., Piana, F., Federico, L., 2010. Geology of the western alps-northern apennine junction area: a regional review. Journal of the Virtual Explorer 36 (9).

Mosca, P., Polino, R., Rogledi, S., Rossi, M., http://doi.org/10.1007/s00531-009-0428-2 2009. New data for the kinematic interpretation of the alps-apennines junction (northwestern italy). International Journal of Earth Sciences 99 (4), 833–849.

Nardi, D., Brachman, R. J., 2003. The description logic handbook. Cambridge University Press, New York, NY, USA, Ch. An Introduction to Description Logics, pp. 1–40.
URL http://dl.acm.org/citation.cfm?id=885746.885748

Narock, T., Szabo, A., Merka, J., 2009. Using semantics to extend the space physics data environment. Computers and Geosciences 35, 791–797.

Piana, F., Fioraso, G., Irace, A., Mosca, P., d'Atri, A., Barale, L., Falletti, P., Monegato, G., Morelli, M., Tallone, S., Vigna, G. B., doi: 10.1080/17445647.2017.1316218 2017. Geology of piemonte region (nw italy, alps–apennines interference zone). Journal of Maps 13 (2).

Polleres, A., Hogan, A., Delbru, R., Umbrich, J., 2013. Rdfs and owl reasoning for linked data. In: Reasoning Web. The 9th International Summer School on Semantic Technologies for Intelligent Data Access. Vol. 8067 of Lecture Notes in Computer Science Series. Springer, pp. 91–149.

Raskin, R., Pan, M., 2005. Semantic web for earth and environmental terminology (sweet - http://ceur-ws.org/vol-83/sia_7.pdf). Computer and Geosciences 31 (9), 1119–1125.

Richard, S., Allison, L., 2016. Capitalizing on global demands for open data access and interoperability – the usgin story. In: EGU2016-9540, Geophysical Research Abstracts,. Vol. 18. http://meetingorganizer.copernicus.org/EGU2016/EGU2016-9540.pdf, EGU General Assembly.

Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., Katz, Y., 2007. Pellet: A practical owl-dl reasoner. Web Semantics: Science, Services and Agents on the World Wide Web 5 (2).

Woodcock, R., Simons, B., Duclaux, G., Cox, S., 2010. Auscope's use of standards to deliver earth resource data,. In: Geophysical Research Abstracts 12, EGU General Assembly. p. 1556.

Zhao, P., Di, L., Yu, G., Yue, P., Wei, Y., Yang, W., 2009. Semantic web-based geospatial knowledge transformation. Computers and Geosciences 35, 798–808.